

Uma abordagem computacional para preenchimento de falhas em dados micro meteorológicos

A computational approach for gap filling in micrometeorological data

RESUMO

Estações micro meteorológicas utilizam equipamentos para captar dados sobre fenômenos climatológicos. Essa captação está sujeita a falhas e influências externas que ocasiona ausência de dados no conjunto de dados gerados. Técnicas matemáticas e computacionais são comumente usadas com o objetivo de preencher essas falhas nos dados. Este trabalho apresenta uma abordagem computacional que utiliza técnica de redes neurais, combinada com algoritmos genéticos, aplicada a dados reais com o objetivo de preencher falhas em séries de temperatura em uma região de cerrado no estado de Mato Grosso. Nos testes realizados, os coeficientes de correlação variaram entre 0,79 e 0,96 e o erro médio absoluto entre de 0,62 e 1,22, mostrando um bom desempenho da rede neural para uma série de dados com valores ausentes.

PALAVRAS-CHAVE: micro meteorologia, fenômenos climatológicos.

ABSTRACT

Micrometeorological stations use equipment to capture data related to climatological phenomena. Such devices are prone to failure, as missing data. Mathematical and computational techniques are commonly used in order to fill the gaps in data set. This paper presents a computational approach that uses neural networks technique, combined with genetic algorithms applied to real data in order to fill gaps in a series of temperature. In tests, the correlation coefficients varied between 0.79 and 0.96 and mean absolute error of between 0.62 and 1.22, showing a good performance of the neural network to a series of data with missing values.

KEYWORDS: micro meteorological, climatological phenomenon.

Thiago Meirelles Ventura

Doutorando pelo Programa de Pós-Graduação em Física Ambiental na UFMT

Cuiabá, MT, Brasil
thiago@ic.ufmt.br

Allan Gonçalves de Oliveira

Doutorando pelo Programa de Pós-Graduação em Física Ambiental na UFMT

Cuiabá, MT, Brasil
allan@ic.ufmt.br

Henrique Oliveira Marques

Mestrando pelo Programa de Pós-Graduação em Ciências de Computação e Matemática Computacional na USP

São Carlos, SP, Brasil
hom@icmc.usp.br

Roberto Silva Oliveira

Graduando em Ciência da Computação na Universidade Federal Mato Grosso (UFMT)
Cuiabá, MT, Brasil
silva.robertoolv@gmail.com

Claudia Aparecida Martins

Doutora em Inteligência Artificial, Professora na UFMT
Cuiabá, MT, Brasil
claudia@ic.ufmt.br

Josiel Maimone de Figueiredo

Doutor em Banco de Dados, Professor na UFMT
Cuiabá, MT, Brasil.
josiel@ic.ufmt.br

Andréia Gentil Bonfante

Doutora em Inteligência Artificial, Professora na UFMT
Cuiabá, MT, Brasil
andreia.bonfante@ic.ufmt.br

INTRODUÇÃO

Discussões sobre o meio ambiente e mudanças climáticas estão tomando cada vez mais importância no cenário mundial. Essa importância levou a necessidade de se entender como se comportam os fenômenos climáticos, objetivando modelar, prever e correlacionar tais fenômenos. Para ter um entendimento melhor sobre esses fenômenos, são utilizadas as estações micrometeorológicas para captar, armazenar e analisar dados desse contexto (SERAFIM JÚNIOR, 2008).

Estações de micrometeorologia utilizam equipamentos específicos para medir as variáveis climáticas. Normalmente, vários equipamentos são instalados no campo para coletar e armazenar dados para análise, e cada equipamento é responsável por mensurar uma ou mais variáveis climáticas, como temperatura, umidade relativa do ar, insolação, entre outras.

Como todo aparelho eletrônico, os equipamentos de medições de micrometeorologia estão sujeitos a falha. Tais falhas podem ser resultado de um erro técnico ou até mesmo de fenômenos naturais atrapalhando a leitura dos dados. Essas falhas comprometem as análises realizadas com base nos dados das estações de micrometeorologia. Por isso é importante que essas falhas sejam detectadas e preenchidas com dados próximos do que seria o real captado, para uma análise mais confiável.

Preencher falhas de dados ausentes consiste em estimar os valores, modelando o comportamento do fenômeno baseado em dados históricos. Fenômenos meteorológicos são complexos e não lineares, pois dependem da interação entre os vários elementos climáticos. Técnicas computacionais como

Redes Neurais Artificiais (RNA) têm sido utilizadas com sucesso para modelar relações envolvendo séries temporais complexas (ZANETTI et al., 2007). A utilização de RNAs em problemas de modelagens complexas se deve em função de sua estrutura não linear e a capacidade de captar características mais complexas dos dados, o que nem sempre é possível com a utilização das técnicas estatísticas tradicionais (GALVÃO et al., 1999).

Trabalhos desenvolvidos relacionados com preenchimento de falhas, podem ser encontrados em Tsukahara (2010), no qual foi utilizada uma RNA para o preenchimento de falhas em séries horárias de dados meteorológicos do estado do Paraná-Brasil. Em Lima (2010), a técnica de Support Vector Machines (SVM) foi utilizada para preenchimento de falhas em dados espaciais binários de Precipitação.

No entanto, quando se trabalha com a técnica de RNA existe uma certa dificuldade de encontrar a melhor estrutura da rede que, geralmente, consiste em investigar todo um espaço de estados possíveis (MATOS NETO et al., 2005). Como a análise de todas as possibilidades é impraticável, é possível utilizar algum algoritmo de busca para encontrar uma solução satisfatória. Nesse contexto, técnicas de Algoritmos Genéticos (AG) podem ser utilizadas para auxiliar na definição da estrutura da rede, por ser um método de busca que tem como característica encontrar uma solução baseada no ótimo global (ASSUMPÇÃO et al., 2011).

Alguns trabalhos já foram desenvolvidos com o objetivo de otimizar a estrutura de uma RNA por meio de AG. Em Matos Neto et al. (2005) pode-se observar a utilização de AG para a seleção de entradas da RNA para previsão de dados em séries temporais. Já em Ahmad et al. (2010), além de selecionar as melhores entradas, foi definido o número de neurônios na camada oculta de uma RNA para diagnóstico

de câncer. Em Ooba et al. (2006) foi definido as entradas, as taxas de aprendizagem e de momentum e, ainda, os pesos iniciais das conexões para corrigir falhas em dados de fluxo de carbono.

Para preenchimento de falhas, além das técnicas computacionais citadas, técnicas estatísticas tradicionais de tratamento de séries temporais também são utilizadas. Em Biudes (2010) foram utilizados modelos de média móvel, exponencial simples e exponencial duplo para preenchimento de falhas em valores de fluxo de seiva obtidos pelo método de balanço de calor no caule, em uma mangabeira sob irrigação e uma não irrigada. Outra alternativa estatística, utilizada em Hui (2004) e em Tatsch (2007), é o método Multiple Imputation, na qual foi usada para preencher falhas em medidas do aparelho de eddy covariance e para dados de fluxo de energia.

Vale ressaltar que nos trabalhos citados, as técnicas são aplicadas de forma totalmente dependentes do tipo de dado utilizado, ou seja, a aplicação se torna bastante específica a fim de tratar uma determinada variável climática, sendo necessário tempo e um conhecimento profundo do domínio sobre os respectivos dados, para atingir as necessidades do mesmo.

Nesse sentido, o objetivo deste trabalho é propor uma abordagem computacional para facilitar o preenchimento de falhas (valores ausentes) em dados climatológicos de maneira eficaz, utilizando técnicas de AG e RNA. Os algoritmos genéticos são utilizados para determinar os melhores parâmetros possíveis da arquitetura de uma rede neural artificial, para que, posteriormente, a mesma possa estimar valores precisos/aproximados visando o preenchimento das falhas. A principal contribuição desse trabalho consiste no desenvolvimento dessa abordagem

Tabela 1. Exemplo de dados obtidos pelos equipamentos micrometeorológicos

Dia	P	T	U	I	V	N	G
73	991,9	27,7	79,8	85	8	11,99	10,42
74	992,3	26,9	84,8	63	7,66	11,98	9,16
75	992,0	26,5	86,5	51	7,33	11,96	8,35
76	992,2	26,1	89,0	26	6,66	11,95	6,88
77	993,0	24,7	98,3	7	5,66	11,93	5,81
77	992,5	25,8	87,0	33	6,66	11,92	7,29
78	994,3	26,0	86,0	40	7	11,90	7,58

considerando que não existe a preocupação com quais variáveis específicas os dados estão relacionados ou quais parâmetros iniciais devem ser atribuídos na estrutura computacional. O desempenho dessa abordagem é mostrado em uma aplicação com dados reais micrometeorológicos.

MATERIAIS E MÉTODOS

Os dados utilizados neste trabalho foram coletados na Estação Agroclimatológica Padre Ricardo Remetter, localizada em uma Fazenda Experimental (latitude: -15,78°; longitude: -56,07°; altitude: 140 metros) da Universidade Federal de Mato Grosso (UFMT). Está situada no município de Santo Antônio de Leverger na Baixada Cuiabana, Mato Grosso, Brasil, distante aproximadamente 28 quilômetros da capital Cuiabá.

Foram utilizados 7 sensores que medem valores de pressão atmosférica, temperatura média, umidade relativa do ar, insolação, velocidade média do vento, fotoperíodo e radiação global. Os sensores estiveram ligados durante 24 horas em todos os dias do ano de 2006, sendo que cada um gerava uma nova medição de 15 em 15 minutos.

Na Tabela 1 é apresentada uma amostra desses dados com o dia (em formato Juliano), as medidas de pressão atmosférica (P), temperatura média (T), umidade relativa do ar (U), insolação (I),

velocidade média do vento (V), fotoperíodo (N) e radiação global (G).

Como pode ser visto, os dados foram agrupados dia a dia. Os valores mostrados na Tabela 1 consistem das médias dos dados coletados em cada dia durante um ano. Logo, tem-se uma base com 365 exemplos de medidas de médias diárias.

Experimentos Realizados

A base de dados é separada em dois conjuntos, um para o treinamento e outro para o teste. A base de treinamento será utilizada para o aprendizado de como as variáveis climáticas se comportam. Para tanto, é fornecida para a estrutura todos os valores que serão preenchidos. Já a base de testes será utilizada para comparar o resultado calculado no experimento com o resultado real, visto que o valor que deveria ser preenchido é conhecido.

Para verificar o desempenho e efetividade da abordagem, a base de dados foi dividida de cinco formas diferentes, para cada conjunto de treinamento e teste, envolvendo diversas quantidades de dados com falhas. Assim, na primeira divisão apenas 5% dos dados foram aleatoriamente selecionados como dados falhos. Isso quer dizer que esses 5% foram utilizados na fase de testes, onde o sistema não tinha conhecimento sobre qual o valor correto para o respectivo valor. Os outros 95%

serviram para a fase de treinamento. A mesma abordagem foi utilizada para as outras divisões, sendo elas de 10%, 20%, 30% e, por fim, 40% representados como dados ausentes.

Os dados foram normalizados para que todas as entradas tivessem a mesma influência sobre o resultado (OH et al., 2007). Dessa forma, todos os valores ficaram no intervalo de [-1,1]. No final do processo, o valor da saída foi desnormalizado e obteve-se o valor estimado da variável em seu formato original. O processo de normalização e desnormalização são descritos nas Eq.(1) e Eq.(2) respectivamente, onde pn é o valor normalizado, p é o valor real, $minp$ e $maxp$ são os valores mínimos e máximos de p .

$$pn = \frac{2 \cdot (p - minp)}{(maxp - minp) - 1} \quad (1)$$

$$p = 0.5 \cdot (pn + 1) \cdot (maxp - minp) + minp$$

Estrutura Proposta

Na Figura 1 é mostrado, resumidamente, um diagrama da abordagem criada, na qual primeiramente é fornecido ao sistema um conjunto de dados coletados pelos diversos sensores dos equipamentos da estação micrometeorológica.

O segundo passo é determinar a arquitetura da RNA

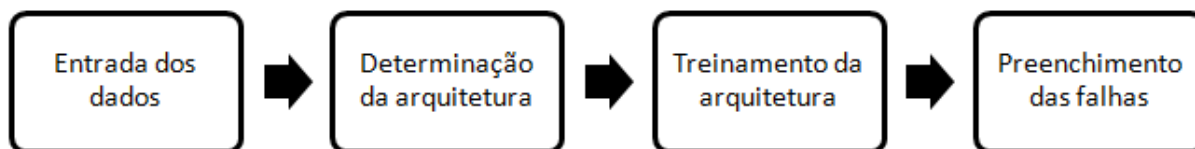


Figura 1. Abordagem para preenchimento de falhas em dados micrometeorológicos.

que será utilizada para preencher as falhas. Para determinar essa arquitetura foi utilizado o AG. Além de o AG testar várias combinações de parâmetros da RNA, ele também avalia quais sensores têm relações aos dados que terão suas falhas preenchidas, ou seja, um sensor não será utilizado caso ele não contribua com as detecções dos padrões nos dados.

Ao final do processamento do AG tem-se a definição da arquitetura da RNA. Assim é possível realizar o treinamento dessa rede para que a mesma seja capaz de realizar os preenchimentos de falhas.

Por fim, o último passo é utilizar a RNA treinada para que os dados com falhas possam ser preenchidos.

Definição do Algoritmo Genético

Algoritmos genéticos foram inicialmente propostos por John Holland (1975), e são baseados no princípio da seleção natural de Charles Darwin, fundamentando-se na afirmação de que os indivíduos mais adaptados têm maior chance de sobreviver e gerar descendentes (LACERDA & CARVALHO, 1999).

Os AGs são amplamente utilizados para problemas de otimização, seu princípio básico consiste em fazer evoluir um conjunto de soluções candidatas iniciais (indivíduos), para uma solução ótima. Ele inicia selecionando aleatoriamente certo número de indivíduos dentro do espaço de busca. Os indivíduos selecionados são avaliados em relação à capacidade de resolver o problema, e essa capacidade é expressa numericamente pela

avaliação do indivíduo, através de uma função objetivo (MICHALEWICZ, 1998). Com base nessa informação, uma nova população é formada através do uso de operadores probabilísticos de seleção, *crossover* (recombinação) e mutação.

O AG neste trabalho tem duas funções importantes:

- a) determinar quais sensores serão utilizados como entrada na RNA e
- b) determinar os parâmetros de maior importância para o treinamento da RNA.

Para preencher os dados ausentes relacionados a um sensor, que representa uma variável climática, são utilizados como fonte de informação os dados existentes dos outros sensores. Entretanto, isso não significa que todos os sensores enviados no conjunto de dados devem ser utilizados. Os dados coletados pelo sensor podem ou não ser influenciados por outra variável climática. Logo, o AG deve fazer uma seleção para determinar quais sensores serão utilizados para preencher a falha de um determinado fenômeno.

Com relação aos parâmetros da RNA, o AG testa valores diferentes para as funções de ativação, algoritmo de treinamento, taxa de aprendizagem e taxa de *momentum*.

A implementação do algoritmo genético bem como da rede neural foi desenvolvida utilizando o ambiente de programação do software Matlab® (MATLAB, 2011). Em resumo, o AG determina:

1. Os sensores que serão utilizados para o

determinado problema.

2. A função de ativação para a camada oculta da RNA: *Linear transfer function*, *Log sigmoid transferfunction*, *Hyperbolic tangent sigmoid transfer function* e *Saturating linear transfer function*.
3. A função de ativação para a camada de saída da RNA: *Linear transfer function*, *Log sigmoid transfer function*, *Hyperbolic tangent sigmoid transfer function* e *Saturating linear transfer function*.
4. O algoritmo de treinamento da RNA: *Gradient descent backpropagation*, *Levenberg-Marquardt backpropagation*, *Resilient backpropagation* e *Gradient descent with momentum and adaptive learning rate backpropagation*.
5. A taxa de aprendizado (*lr*) da RNA.
6. A taxa de *momentum* (*mc*) da RNA.

Os indivíduos do AG responsáveis por realizar essa tarefa foram compostos por 18 bits, conforme representado na Figura 2. Cada indivíduo possui dois bits que foram reservados para a função de ativação para a camada oculta (**a**), possibilitando que quatro tipos de funções fossem escolhidas. Dois bits para a função de ativação para a camada de saída (**b**) e dois para o algoritmo de treinamento (**c**). Para

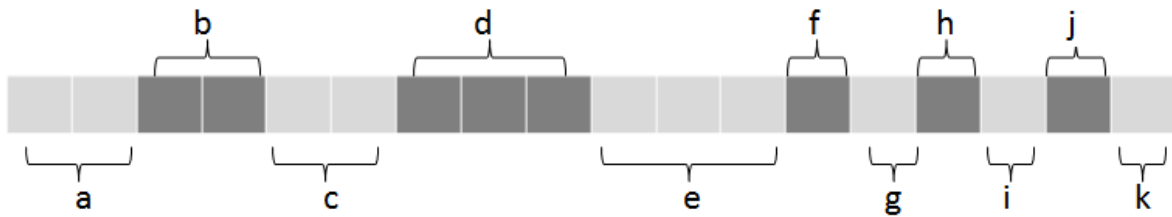


Figura 3. Representação de um indivíduo do AG.

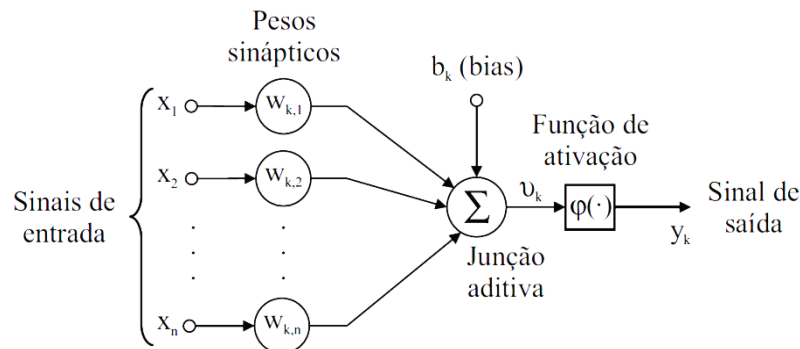


Figura 2. Estrutura de uma Rede Neural (HAYKIN, 2001).

as taxas de aprendizado (**d**) e de *momentum* (**e**) foram reservados 3 bits para cada, possibilitando que os seus valores variem de 0,1 a 0,8. E para cada sensor foi reservado 1 bit indicando se o mesmo deve ou não ser utilizado (**f**, **g**, **h**, **i**, **j**, **k**).

A função objetivo do AG foi constituída por uma RNA utilizando os parâmetros definidos no indivíduo atual e processá-la para verificar como foi o seu desempenho. Quando a população de indivíduos consegue um erro aceitável retornado pela RNA em seu treinamento, o AG termina e o resultado é apresentado mostrando as melhores combinações de entradas e parâmetros para a rede neural que tem a função de preencher as falhas dos dados captados pelos sensores.

Definição da Rede Neural

As redes neurais artificiais são técnicas computacionais cujo funcionamento baseia-se na estrutura neural de seres vivos e que realizam tarefas de computação adquirindo conhecimento através de experiência, construída por um processo de aprendizagem, tendo

como principais vantagens as características de adaptabilidade, generalização e tolerância a falhas (HAYKIN, 2001).

Todas as redes testadas neste trabalho foram do tipo perceptron de múltiplas camadas (MLP, *multilayer perceptron*). Nesse tipo de rede o sinal de entrada se propaga para frente (*feedforward*), camada por camada, sendo em seguida retropropagado para a correção do erro (ajuste dos pesos sinápticos). Este procedimento é repetido durante várias iterações até a finalização do treinamento (ZANETTI, 2008).

O funcionamento de uma rede neural (Figura 3) é basicamente uma função dos sinais de entrada pelos seus respectivos pesos sinápticos ($W_{k,j}$). O *bias* funciona aumentando ou diminuindo a influência do valor da entrada líquida para a ativação do neurônio; já a função de ativação funciona restringindo a amplitude de saída de determinado neurônio e adicionando não-linearidade ao modelo (ZANETTI, 2008).

Neste trabalho foram utilizadas duas redes neurais, uma

para servir de função objetivo para o AG e outra para fazer a etapa final do preenchimento de falhas, sendo que ambas foram implementadas no Matlab.

As duas redes são semelhantes em relação à sua estrutura. A diferença principal é que os valores dos pesos de entrada, dos pesos da camada oculta e do limiar são pré-definidos baseando-se nos melhores resultados quando a rede neural da função objetivo foi executada.

Em ambas as redes, o número de camadas e neurônios foi estruturado baseando-se no Teorema de Kolmogorov-Nielsen, apresentado por Kovacz (1997), onde:

"Dada uma função contínua arbitrária

$$f: [0,1]^m \rightarrow R^m, f(x) = y,$$

existe sempre, para f uma implementação exata com uma rede neural de três camadas, sendo a camada de entrada um vetor de dimensão m , a camada oculta composta por $2n + 1$ neurônios, e a camada de saída com M

neurônios, representando os M componentes do vetor y .”

De acordo com o teorema de Kolmogorov-Nielsen, n representa o número de neurônios na camada de entrada, assim, para a RNA utilizada na abordagem proposta tem-se no máximo $n = 6$, o que significa que contém no máximo $2n + 1 = 13$ neurônios na camada oculta. Esse valor pode ser menor já que o número de entradas (igual ao número de sensores) pode ser diminuído caso seja detectado que não há necessidade do uso de algum sensor.

A validação estatística da estimativa do modelo foi baseada nos erros individuais e_i de cada estimativa Eq.(1), onde $(P_i; i = 1,2,\dots,n)$ são os valores estimados e $(O_i; i = 1,2,\dots,n)$ são os valores reais (WILLMOTT & MATSUURA, 2005). Com o erro individual de cada estimativa, é calculado o desempenho do modelo através do Erro Médio Absoluto (*Mean Absolute Error - MAE*) Eq.(2) que, de acordo com Willmott et al. (2009), é a melhor forma de avaliação para modelos ambientais, devido principalmente a presença de *outliers* e dados com desvio de normalidade.

$$e_i = P_i - O_i \quad (3)$$

$$MAE = \left[n^{-1} \sum_{i=1}^{i=n} |e_i|^2 \right] \quad (4)$$

RESULTADOS E DISCUSSÕES

Normalmente, nos trabalhos citados anteriormente de preenchimento de falhas, os procedimentos são realizados para uma determinada variável climática. Entretanto, nesta abordagem, uma das principais vantagens é a possibilidade de preenchimento de falhas de diversas variáveis climáticas. Isso significa que não é necessário haver um pré-processamento ou configuração diferenciada para cada variável climática que será tratada. A configuração mais adequada, para realizar o procedimento de preenchimento de falhas, será determinada automaticamente na execução da estrutura proposta.

Para ilustrar o desempenho dessa proposta, dos sete sensores existentes na base de dados, o sensor de temperatura do ar foi escolhido para simular o seu preenchimento usando dados dos outros sensores. Da mesma forma que a temperatura foi escolhida para demonstrar os resultados, outra variável climática poderia ser escolhida sem nenhum problema, ou seja, o procedimento seria exatamente o mesmo. A única exigência neste caso é que os dados captados pelos outros sensores estejam disponíveis na mesma base de dados que estão relacionados à variável que será tratada.

Na Tabela 2, como primeira etapa da proposta, são mostrados os melhores resultados encontrados pelo AG para cada teste realizado para o sensor de temperatura do ar, envolvendo as cinco simulações realizadas, cujas diferenças estão na quantidade de falhas na base de

dados (de 5% a 40%).

Onde f_1 é a função de ativação utilizada para a camada oculta, f_2 é a função de ativação utilizada para a camada de saída, *train* o algoritmo de treinamento, *lr* a taxa de aprendizado, *mc* a taxa de momentum e de s_1 a s_6 simboliza se o respectivo sensor foi ativado ou não.

As funções de ativações obtidas foram a *Saturating linear transfer function (satlin)*, *Linear transfer function (purelin)* e a *Hyperbolic tangente sigmoid transfer function (tansig)*. Os algoritmos de treinamentos encontrados foram o *Resilient backpropagation (trainrp)*, *Gradient descent back-propagation (traingd)* e o *Levenberg-Marquardt back-propagation (trainlm)*.

As colunas s_1, s_2, s_3, s_4, s_5 e s_6 representam respectivamente os seguintes sensores: pressão atmosférica, umidade relativa do ar, insolação, velocidade média do vento, fotoperíodo e radiação global.

Os coeficientes de correlação entre os valores medidos e os valores estimados pela rede são apresentados na Figura 4. Como esperado, os valores mostram que há forte correlação entre os valores medidos e estimados, sendo que a maior correlação encontrada foi para o teste com 5% de falhas ($R = 0,96031$) e o pior para o teste com 40% de falhas ($R = 0,79396$).

O desempenho ainda é avaliado usando a análise da evolução do MAE para cada teste, como é apresentado na Figura 5. Os valores de MAE indicam resultados satisfatórios para os testes com 5% e

Tabela 2. Resultados encontrados pelo AG nos testes realizados.

% Falhas	f_1	f_2	train	lr	mc	s_1	s_2	s_3	s_4	s_5	s_6
5	tansig	purelin	trainlm	0.3	0.8		sim		sim	sim	sim
10	tansig	purelin	trainrp	0.2	0.3		sim	sim	sim	sim	sim
20	tansig	purelin	trainlm	0.3	0.8	sim	sim	sim			sim
30	satlin	purelin	trainrp	0.2	0.3		sim		sim	sim	sim
40	tansig	tansig	traingd	0.8	0.4		sim		sim	sim	sim

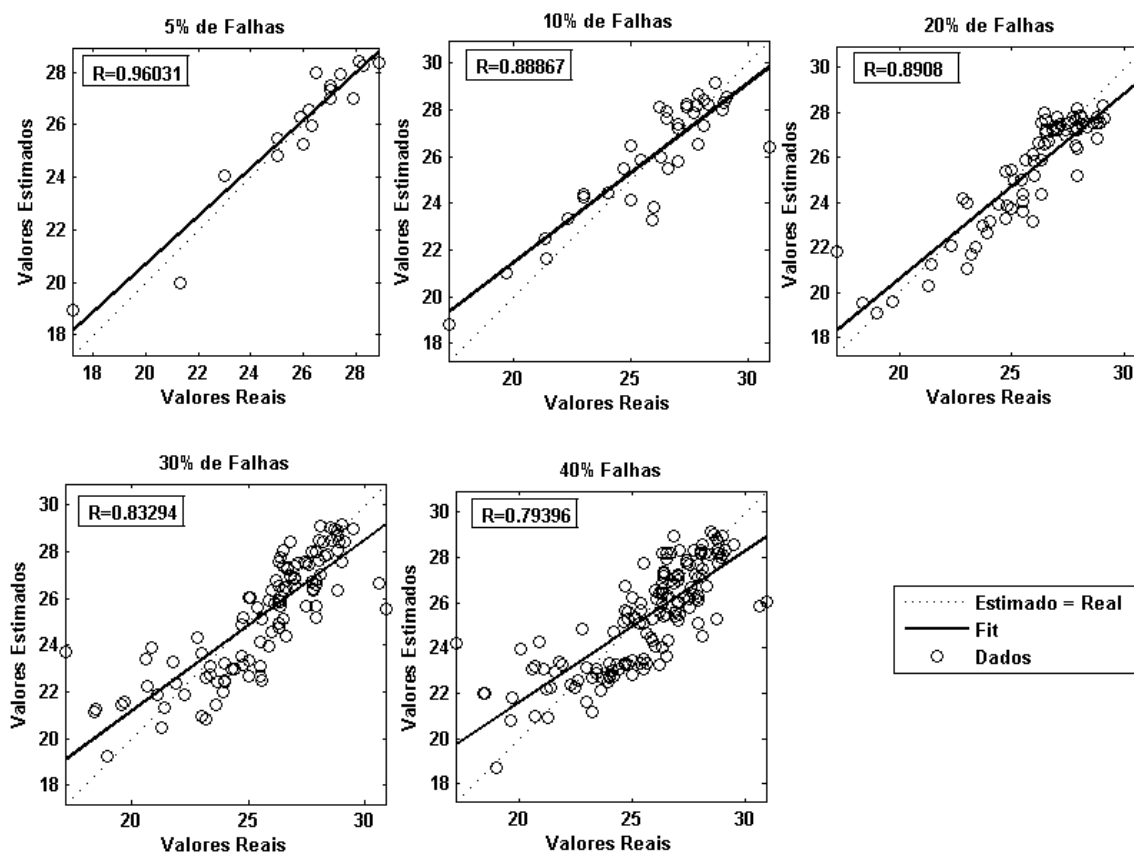


Figura 5. Valores dos coeficientes de correlação para cada teste.

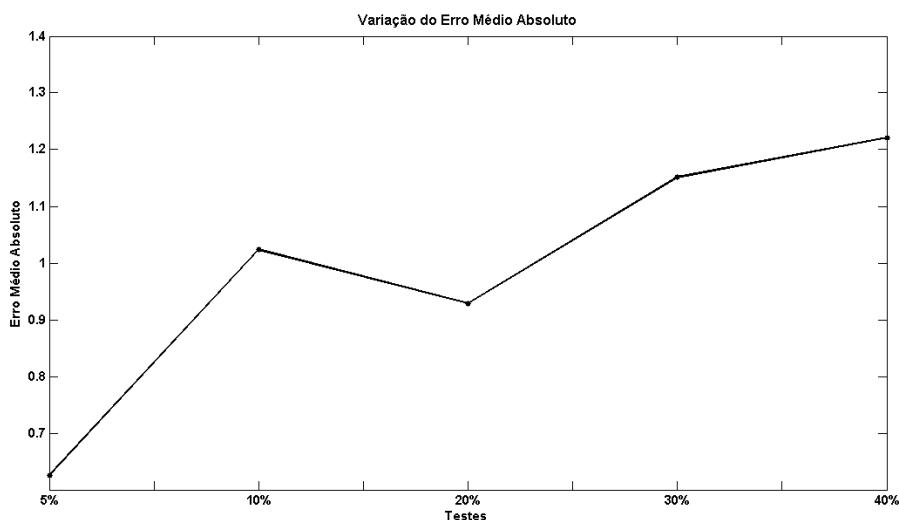


Figura 4. Evolução do erro médio absoluto

20% de falhas. O erro relativamente alto do teste com 10% de falhas em relação a 5% e 20% provavelmente está relacionado pela escolha aleatória dos dados para teste, no qual dados menos representativos podem ter sido escolhido para o treinamento da rede. Já os valores

de MAE encontrados para os testes com 30% e 40% indicam a limitação da rede ao aumentar o número de falhas e consequentemente a diminuição do conjunto de treinamento.

O gráfico mostrado na Figura 6 apresenta um comparativo

entre o valor real medido pelo sensor de temperatura (sem falhas) com os valores obtidos usando a abordagem proposta neste trabalho, para os testes realizados. Pelo gráfico da Figura 6, pode ser observado que com apenas 5% de falhas na série de dados, o

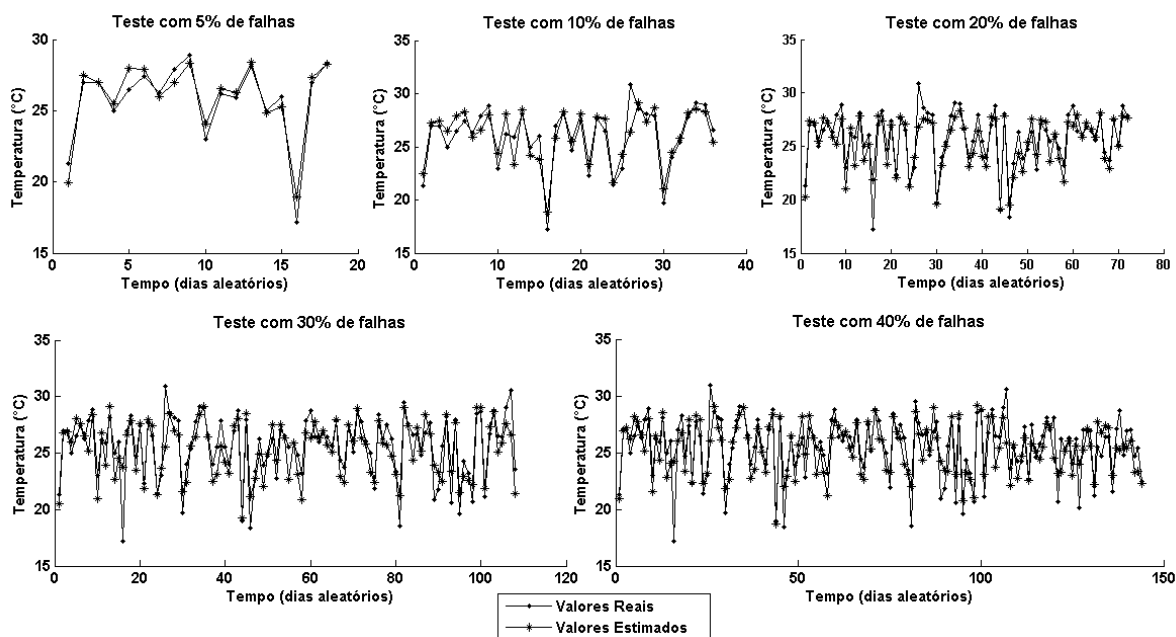


Figura 6. Valores comparativos entre os valores esperados e os valores obtidos pela rede.

comportamento da estimativa dos valores obtidos é muito eficiente. Com o aumento da quantidade de falhas na série de dados essa eficiência diminui, o que já era esperado. Mas é importante observar que mesmo com 40% dos dados faltantes a abordagem teve um desempenho satisfatório no preenchimento de tais falhas.

CONSIDERAÇÕES FINAIS

Neste trabalho foi apresentado uma abordagem para que as falhas (valores ausentes) em séries de dados de estações micrometeorológicas fossem preenchidas por valores aproximados baseando-se nas medidas de outras séries de dados obtidas no mesmo local e momento da falha detectada.

Analisando os resultados obtidos entende-se que é possível simular e preencher dados reais micrometeorológicos usando a abordagem definida, na qual utiliza algoritmos genéticos para estimar os parâmetros da rede neural e redes neurais para estimar valores ausentes de sensores. Os resultados foram avaliados usando o coeficiente de correlação e o erro

médio absoluto, sendo que no teste com a menor quantidade de falhas foi obtido respectivamente 0,96 e 0,62. Os resultados obtidos foram considerados bons pelo especialista do domínio para corrigir séries de dados com pequenos erros, mas também pode ser considerada eficiente quando a série de dados apresenta uma grande parte de falhas.

Geralmente, os dados obtidos das estações micrometeorológicas são fonte de informação de vários trabalhos científicos envolvendo o meio ambiente. Com o preenchimento das falhas que normalmente aparecem nos dados micrometeorológicos, esta abordagem pode ser uma ferramenta importante para beneficiar pesquisas da área.

Alguns trabalhos, como destacado em Guo & Uhrig (1992) e em Vafaie & Imam (1994), consideram que a utilização do AG na seleção de entradas para uma RNA possui esforço computacional muito grande, pois cada indivíduo numa população representa uma possível seleção de entradas para a rede, a qual necessita de ser testada. Entretanto, mesmo

considerando este fato, é importante destacar que, para o analista dos dados, ou o usuário final, a automatização desta fase auxilia de forma robusta e prática todo o procedimento, já que esta atividade geralmente demanda muito esforço manual. No entanto, quando o tempo do processamento é realmente prioridade, é possível utilizar técnicas de paralelismo para reduzir o tempo de processamento do AG e, assim, melhorar o desempenho computacional, como apresentado em Kattan et al. (2010) e em Falahiazar et al. (2012).

Como trabalhos futuros, novos experimentos estão sendo realizados para verificar melhor o desempenho utilizando outros tipos de sensores e uma quantidade maior de dados.

AGRADECIMENTOS

Os autores agradecem à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo apoio financeiro, ao Programa de Pós Graduação em Física Ambiental (PPGFA) e ao Instituto de

Computação (IC) da Universidade Federal de Mato Grosso pelo incentivo a pesquisa.

REFERÊNCIAS

- AHMAD, F.; MATISA, N. A.; HUSSAIN, Z.; BOUVILLE, R.; OSMAN, M. K. Genetic Algorithm – Artificial Neural Network (GA-ANN) Hybrid Intelligence for Cancer Diagnosis, In **Second International Conference on Computacional Intelligence, Communication System and Networks**, 2010.
- ASSUMPCÃO, M. E.; CHRISTOFORO, A. L.; PANZERA, T. H.; BRITO, J. N.; LAMIM-FILHO, P. C. M. Emprego de Algoritmos Genéticos na Determinação das Dimensões Ótimas de uma Torre Treliçada, In **XI Congresso Nacional de Engenharia Mecânica, Metalúrgica e Industrial**, Porto Alegre - RS, 2011.
- BIUDES, M. S.; CAMPELO JÚNIOR, J. H.; ESPINOSA, M. M.; NOGUEIRA, J. S. Uso de séries temporais em análise de fluxo de seiva de mangabeira. **Ciência e Natura**, Santa Maria, vol. 31, n. 1, p. 65-77, 2009.
- FALAHIAZAR, L.; TESHNEHLABn M.; FALAHIAZAR, A. Parallel Genetic Algorithm Based on a New Migration Strategy, In **International Conference on Recent Advances in Computing and Software Systems**, 2012.
- GALVÃO, C. O.; VALENÇA, M. J. S.; VIEIRA, V. P. P. B.; DINIZ, L. S.; LACERDA, E. G. M.; Carvalho, A. C. P. L. F.; Ludermir, T. B. **Sistemas inteligentes: Aplicações a recursos hídricos e ciências ambientais**. Porto Alegre: UFRGS/ABRH, 1999, 246p.
- GUO, Z.; UHRIG, R. E. Using Genetic Algorithms to Select Inputs for Neural Networks, In **IEEE International Conference**, 1992.
- HAYKIN, S. **Redes Neurais: Princípios e Práticas**. 2ed. Porto Alegre: Bookman, 2001.
- HUI, D.; WAN, S.; SU, B.; KATUL, G.; MONSON, R.; LUO, Y. Gap-filling missing data in eddy covariance measurements using multiple imputation (MI) for annual estimations. **Agricultural and Forest Meteorology**, vol. 121, p. 93-111, jan.2004.
- KATTAN, A. R. M.; ABDULLAH, R.; SALAM, R. A. Training Feed-Forward Neural Networks using a Parallel Genetic Algorithm with the Best Must Survive Strategy, In **International Conference on Intelligent Systems, Modelling and Simulation**, 2010.
- KOVÁCS, Z. L. **Redes Neurais Artificiais: Fundamentos e Aplicações**. Livraria da física, 1997, 107p.
- LACERDA, Estéfane George Macedo de, CARVALHO, Andre Carlos Ponce de Leon Ferreira de.: **Introdução aos algoritmos genéticos**. In: GALVÃO, Carlos Oliveira, VALENÇA, M. J. S.: **Sistemas inteligentes: aplicações a recursos hídricos e ciências ambientais**, Porto Alegre: Ed. Universidade/UFRGS: Associação Brasileira de Recursos Hídricos, p. 99-150. 1999.
- LIMA, C. H. R. **Preenchimento de Falhas em Dados Espaciais Binários de Precipitação Utilizando Máquinas de Vetor de Suporte (Support Vector Machines)**. In: SIMPÓSIO NACIONAL DE PROBABILIDADE E ESTATÍSTICA, 19, 2010, São Pedro.
- MATLAB. The Language Of Technical Computing. Disponível em: <<http://www.mathworks.com/products/matlab/>>. Acesso em: 10 abr. 2011.
- MATOS NETO, P. S. G.; PETRY, G. G.; ATAIDE, J. P. M.; FERREIRA, T. A. E. Combinação de Redes Neurais Artificiais com Algoritmo Genético Modificado para a Previsão de Séries Temporais. In **XXV Congresso da Sociedade Brasileira de**
- Computação**, São Leopoldo - RS, 2005.
- MICHALEWICZ, Z. **Genetic Algorithms + Data Structures = Evolution Programs**. New York : Springer-Verlag,1994.
- OH, C.; Z.AK1, S. H.; MIRZAEI, H.; BUCK, C.; REGNIER, F. E.; ZHANG, X. Neural network prediction of peptide separation in strong anion exchange chromatography. **Bioinformatics**, vol. 23, n. 1, p 114-118, 2007.
- OOBA, M.; HIRANO, T.; MOGAMI, J. I.; HIRATA, R.; FUJINUMA, Y. Comparisons of Gap-filling Methods for Carbon Flux Dataset: A Combination of a Genetic Algorithm and an Artificial Neural Network. **Ecological Modelling**, v. 198, n. 3-4, p. 473–486, 2006.
- SERAFINI JÚNIOR, S. Histórico de Instalação das Estações Meteorológicas do INMET no Estado de Minas Gerais. In **Simpósio de Pós-Graduação em Geografia do Estado de São Paulo**, UNESP - Rio Claro - SP, 2008.
- TATSCH, J.; DA ROCHA, H.; CABRAL, O.; FREITAS, H.; LLOPART, M.; ACOSTA, R.; LIGO, M. Avaliação do método de Multiple Imputation no preenchimento de falhas de fluxos de energia sobre uma área de cana-de-açúcar. **Ciência e Natura**, v. Esp., p. 109-112, 2007.
- TSUKAHARA, R. Y.; JESEN, T.; CARAMORI, P. Utilização de Redes Neurais Artificiais para Preenchimento de Falhas em Séries Horárias de Dados Meteorológicos. **Congresso Brasileiro de Meteorologia**, 2010.
- VAFAIE, H.; IMAM, I. F. Feature Selection Methods: Genetic Algorithms vs. Greedy-like Search, In **International Conference on Fuzzy and Intelligent Control Systems**, 1994.

WILLMOTT, C.; MATSUURA, K.
Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. **Climate Research** **30**: p. 79-82, 2005.

WILLMOTT, C. J.; MATSUURA, K.; SCOTT, M.; ROBESON, S. Ambiguities inherent in sums-of-squares-based error statistics. **Atmospheric Environment**. Vol. 43: p. 749-752, 2009.

ZANETTI, S. S.; SOUZA E. F.; DE CARVALHO, D. F.; BERNARDO, S. Estimação da evapotranspiração de referência no Estado do Rio de Janeiro usando redes neurais artificiais. **Revista Brasileira de Engenharia Agrícola e Ambiental**. Vol. 12, n.2, p.174–180, 2008.

ZANETTI, S. S.; SOUZA, E. F.; OLIVEIRA, V. P. S.; ALMEIDA, F. T., BERNARDO, S. Estimating evapotranspiration using artificial neural network and minimum climatological data. **Journal of Irrigation and Drainage Engineering**. Vol. 33(2): p. 83-89, 2007.

Recebido em: jul/2011
Aprovado em: out/2013